Asymptotic near-efficiency of a "Gibbs-energy" estimating function approach for fitting Matérn covariance models to a dense (noisy) series.

D.A. Girard

CNRS and University J. Fourier, Grenoble

Abstract

Let us call "Gibbs energy" the quadratic form occurring in the maximum likelihood (ML) criterion when fitting a zero-mean multidimensional Gaussian distribution to one realization. We consider a continuous-time Gaussian process Z which belongs to the Matérn family with known regularity index $\nu \geq 1/2$. For estimating the range and the variance of Z from noisy observations on a dense regular grid, we propose two simple estimating functions based on the conditional Gibbs energy mean (CGEM) and the empirical variance (EV). We show that the ratio of the large sample mean square error of the CGEM-EV estimate of the range-parameter over the one of its ML estimate, and the analog ratio for the variance-parameter, both converge (when the grid-step tends to 0) toward a constant, only function of ν, surprisingly close to 1 provided ν is not too large. Extensions of this approach, which may enjoy a very easy numerical implementation, are briefly discussed.

1. Introduction

We consider time-series of length n obtained by observing on a dense regular grid, a continuous-time process Z which is Gaussian, has mean zero and an autocorrelation function which belongs to the Matérn family with regularity index $\nu \geq 1/2$. This family is commonly used, e.g. in geostatistics (see Stein 1999); recall that $\nu = 1/2$ correspond to the well known exponential autocorrelation (in other words, Z is a stationary Ornstein-Uhlenbeck (O.U.) process). The general definition of Matérn processes can be easily formulated in terms of the Fourier transform of their autocorrelation function, namely the spectral density over $(-\infty, +\infty)$:

$$f^*_{\nu,b,\theta}(\omega) = b\, g^*_{\nu,\theta}(\omega), \text{ where } g^*_{\nu,\theta}(\omega) = \frac{c_\nu\, \theta^{2\nu}}{(\theta^2+\omega^2)^{\frac{1}{2}+\nu}}. \tag{1.1}$$

In this paper the constant $c_\nu = \frac{\Gamma(\nu+\frac{1}{2})}{\sqrt{\pi}\Gamma(\nu)}$ is chosen so that $\int_{-\infty}^{\infty} g^*_{\nu,\theta}(\omega)d\omega = 1$. Thus b is the variance of $Z(t)$ and θ is the so-called "inverse-range parameter" (in fact, it is $2\nu^{1/2}/\theta$ which can be interpreted as the effective range or "correlation length", cf Stein (1999, Section 2.10); we will often drop the term "inverse".

We are concerned here with dense grid in the sense that the interval δ between two succesive observations is small relatively to $1/\theta$. The considered processes being mean square continuous, this means that two successive ideal (i.e. without measurement error) observations are "strongly" correlated. See Stein (1999, Chapter 3), for such a setting which shows that standard large-n asymptotic analysis followed by a (less standard) small-δ analysis yields useful insights and good approximations for various real (finite-size) problems.

In this article, we assume that ν is known and we mainly study settings where there are Gaussian i.i.d. measurement errors, or, equivalently for the parametric inference point of view we take here, there is a so-called nugget effect. Notice that introducing such a nugget effect is actually not too restrictive since, of course, data always have a finite number of digits, so one may easily accept to include, in the parametric model, a known nugget equal to the squared finite precision; furthermore this is known to often remedy to numerical instabilities due to ill-conditionning of the linear systems we may encounter in the no-nugget case. However, in this first study, we restrict ourselves to the case where the suspected measurement errors have a known variance. We thus assume that (possibly after appropriate rescaling) the observations is a vector $\mathbf{y}$ of size n whose conditional law given Z is

$$\mathbf{y}|\mathbf{z} \sim N(\mathbf{z}, I) \text{ where } \mathbf{z} \sim N(0, b_0 R_{\theta_0}),$$

with I denoting the identity matrix and R_θ the autocorrelation matrix of the column vector $\mathbf{z} = (Z(\delta), Z(2\delta), \ldots, Z(n\delta))^T$, i.e. the Toeplitz matrix with coefficients given by

$$[R_\theta]_{j,k} = K_{\nu,\theta}(\delta|j-k|),\ j,k = 1,\cdots,n, \text{ where } K_{\nu,\theta}(t) = \int_{-\infty}^{\infty} g^*_{\nu,\theta}(\omega)e^{i\omega t}d\omega.$$

Recall that expressions in terms of Bessel functions are well known for $K_{\nu,\theta}$ (see e.g. Rasmussen and Williams (2006) for explicit very simple expressions for $\nu = 3/2$ and $5/2$).

It is often of great interest to be able to "effectively reduce" the number of parameters, especially when computing the likelihood function is costly. This is classically done in the no-nugget case (i.e. $bR_\theta + I$ is replaced by bR_θ, and $\mathbf{y}$ by $\mathbf{z}$,

in the likelihood function) by noticing that the maximizer b of the likelihood for any fixed θ is simply:

$$\hat{b}(\theta) = (1/n)\mathbf{z}^T {R_\theta}^{-1}\mathbf{z}. \tag{1.2}$$

See Zhang (2004) also for numerical improvements produced by such a “concentration” of the likelihood. Zhang and Zimmerman (2007) recently proposed to use the classical weigthed least square method (not statistically efficient but much less costly than maximum likelihood (ML)) to estimate the range parameters, next, to plug-in these parameters (the θ here) in the likelihood which is then maximized with respect to b (the solution, say $\hat{b}_{\mathrm{ML}}(\theta)$, being the explicit (1.2) in the no-nugget case, obtained iteratively by Fisher scoring otherwise). The idea underlying this method is that, at least for the infill asymptotics context, even if θ is fixed at a wrong value θ_1, the product $\hat{b}_{\mathrm{ML}}(\theta_1)\theta_1^{2\nu}$ still remains an efficient estimator of $b_0\theta_0^{2\nu}$ (see Du et al. (2009) for recent results of this type).

The method we propose here, firstly reverses the roles of variance and range, in that it is based on a very simple estimate for the variance, namely the empirical variance in the no-nugget case, and its corrected version for biais otherwise, which is simply defined by $\hat{b}_{\mathrm{EV}} := n^{-1}\mathbf{y}^T\mathbf{y} - 1$. Secondly we propose to replace the maximization of the likelihood w.r.t. θ by the very simple following estimating equation in θ, in the with-nugget case: solve, with b fixed at $\hat{b}_{\mathrm{EV}}$

$$\mathbf{y}^T A_{b,\theta}\left(I - A_{b,\theta}\right)\mathbf{y} = \mathrm{tr}A_{b,\theta} \text{ where } A_{b,\theta} = bR_\theta\left(I + bR_\theta\right)^{-1}. \tag{1.3}$$

In the no-nugget case, this equation in θ is simply replaced by $\mathbf{z}^T {R_\theta}^{-1}\mathbf{z} = n\hat{b}_{\mathrm{EV}}$. One may call “Gibbs energy” (GE) of the underlying process the quantity $\mathbf{z}^T {R_\theta}^{-1}\mathbf{z}$ and it is easily seen that $b\left(\mathbf{y}^T A_{b,\theta}\left(I - A_{b,\theta}\right)\mathbf{y} + \mathrm{tr}(I - A_{b,\theta})\right)$ is the conditional Gibbs energy mean (CGEM) obtained by taking the expectation of $\mathbf{z}^T {R_\theta}^{-1}\mathbf{z}$, conditional on $\mathbf{y}$, for the candidate parameters b, θ. So equation (1.3) in θ will be called the CGEM-EV estimating equation (GE-EV in the no-nugget case) and we will denote by $\hat{\theta}_{\mathrm{GEEV}}$ this new range parameter estimate.

Notice that it is easily checked that equation (1.3) is equivalent to equate to 0 the derivative w.r.t. b (and not w.r.t. θ!) of the likelihood. This proposal is based on the following two ideas. First, since it is quite plausible that the above idea, underlying Zang and Zimmerman (2007)'s proposal, remains true for a random θ, then, instead of “fixing” θ at θ_1, one may as well adjust θ so that $\hat{b}_{\mathrm{ML}}(\theta)$ coincides with a given value b_1 for the variance; and, denoting $\hat{\theta}_1$ the so-obtained θ (thus $(b_1, \hat{\theta}_1)$ must cancel the b-derivative of the likelihood), the product $b_1\hat{\theta}_1^{2\nu}$ will plausibly be an efficient estimator of $b_0\theta_0^{2\nu}$. Second, it is known that, at least in the no-nugget case, the moment $\sum Z(\delta j)^2$ yields a successfull estimating-equation in the case $\nu = 1/2$ to estimate θ_0 when $b_0\theta_0^{2\nu}$ is known (see Kessler (2000)). So it is tempting to select $b_1 := \hat{b}_{\mathrm{EV}}$, the emprical variance estimate of b_0. The second point is an admittedly weak justification of this method. However,

notice that Yadrenko (Chapter 4 Section 3) has studied, for quite general, mean square continous, stationnary isotropic random fields the properties of a continuous version of the empirical variance in the periodic case, i.e. the Lebesgue integral over sphere of $Z(t)^2$, and he has listed appealing properties of such estimates of the variance of the field. The theorical results we give in this article, will provide in our context, a much stronger, and rather unexpected, justification, for ν not too large (which is the case in numerous applications, see Stein (1999), Rasmussen et al. (2006)).

Of course, in our time-series setting, there now exist rather good implementations of ML; see Chen et al. (2006) for the case when the correlation is rather strong. However, our objective here is to provide first insights into the capability of this method for more computationally complex settings. Indeed, this approach is not limited to observations on one dimensional lattice, and is potentialy not limited to regular grid (a weighted version, with Rieman-sum type coefficients, of the empirical variance shoud then be used instead) and to homogenous measurement errors: some successfull applications for two-dimensional Matérn random fields, using a randomized-trace approximation to $\mathrm{tr}A_{b,\theta}$, are described in Girard (2009), along with some theoretical properties. This approach might be, in principle, applied to other two-parameters models than the Matérn family (of course it is presently restricted to scalar θ). Encouraging experimental results are obtained in Girard (2009) for the commonly used spherical autocovariance functions.

The rest of this article is structured as follows. Section 2 lists additional notations. In the asymptotic framework we adopt here (which may be thought as intermediate between the infill and increasing-domain frameworks) we first show, in Section 3, that, even when b is arbitrarily fixed to b_1, the CGEM estimating equation is a quite well-behaved estimating equation: it converges toward a monotonic equation whose the root θ_1 satisfies $b_1\theta_1^{2\nu} = b_0\theta_0^{2\nu}$. Next we study in Section 4 what is sacrificed when plugging-in the simple $\hat{b}_{\mathrm{EV}}$ in the CGEM estimating equation, as compared to ML. The performance-loss is classically quantified by the asymptotic mean square errors of the two components of the CGEM-EV estimate as related to those of the ML estimate. We show that what is sacrificed, which depends on ν, is quite reasonable provided ν is not too large, independently of b_0 and θ_0. Indeed asymptotic efficiency is reached as ν decreases to $1/2$. Proofs are outlined Section 5.

2. Further notations

Of course "time" could be replaced everywhere by "space": we choose this vocabulary since we use in several places of the paper the classical time-series theory. Thus we assume everywhere without loss of generality that Z_δ defined by $Z_\delta(i) = Z(\delta i)$ is observed at times $i = 1, 2, \cdots, n$. The spectral density on $(-\pi, \pi]$

of Z_δ is thus $f^\delta_{\nu,b,\theta} = b\, g^\delta_{\nu,\theta}$ with

$$g^\delta_{\nu,\theta}(\cdot) := \frac{1}{\delta}\sum_{k=-\infty}^{\infty} g^*_{\nu,\theta}\left(\frac{\cdot + 2k\pi}{\delta}\right) = \sum_{k=-\infty}^{\infty} g^*_{\nu,\alpha}\left(\cdot + 2k\pi\right) \quad \text{where} \quad \alpha = \delta\theta,$$

the equality between the two sums resulting from the particular "variance-scale" parameterisation of the Matérn family. We will also denote simply by $g_{\nu,\alpha}(\cdot)$ the second sum (α is the range-parameter for the series on $\mathbb{Z}$). Simple closed expressions for $g^\delta_{\nu,\theta}$ are available only when $\nu - 1/2$ is a small integer, says $0, 1$ or 2 : they then coincide with particular constrained ARMA spectral densities.

Let us define the "Bayesian filter" $a^\delta_{b,\theta}(\cdot)$ and its "un-aliased" version $a^*_{b,\alpha}(\cdot)$, for $\alpha = \delta\theta$, by

$$a^\delta_{b,\theta}(\cdot) := \frac{b g^\delta_{\nu,\theta}(\cdot)}{b g^\delta_{\nu,\theta}(\cdot) + (2\pi)^{-1}}, \quad a^*_{b,\alpha}(\cdot) := \frac{b g^*_{\nu,\alpha}(\cdot)}{b g^*_{\nu,\alpha}(\cdot) + (2\pi)^{-1}}.$$

As is well known $a^\delta_{b,\theta}$ is the spectral characteristic of the "$n = +\infty$" version of $A_{b,\theta}$ (a convolution of series defined on $\mathbb{Z}$) so that, e.g., $\lim_{n\to+\infty} n^{-1}\mathrm{tr}A_{b,\theta} = (2\pi)^{-1}\int_{-\pi}^{\pi} a^\delta_{b,\theta}(\lambda)\mathrm{d}\lambda$ for any fixed δ, b, θ.

A function which will play an important role in this article is the derivative of $\log(g^\delta_{\nu,\theta}(\cdot))$ w.r.t. θ, and its un-aliased approximation (up to a factor δ):

$$h^\delta_{\nu,\theta} := \partial \log(g^\delta_{\nu,\theta})/\partial\theta, \quad h^*_{\nu,\alpha} := \partial \log(g^*_{\nu,\alpha})/\partial\alpha.$$

For any $f : [-\pi, \pi] \to \mathbb{R}$, s.t. $\int_{-\pi}^{\pi} [a^\delta_{b,\theta}(\lambda)]^2 f(\lambda)\mathrm{d}\lambda \neq 0$, we define the weighted coefficient of variation of f by

$$J_{\delta,b,\theta}(f) \;:=\; \frac{\frac{1}{\int w}\int w\left[f - \left(\frac{1}{\int w}\int wf\right)\right]^2}{\left(\frac{1}{\int w}\int wf\right)^2} = \frac{\frac{1}{\int w}\int wf^2}{\left(\frac{1}{\int w}\int wf\right)^2} - 1, \text{ where } w := (a^\delta_{b,\theta})^2.$$

Above and throughout this paper, "$\int$" will denote integrals over $[-\pi, \pi]$. We will also use the notation g_0 (resp. h_0) for the function $g^\delta_{\nu,\theta}$ (resp. $h^\delta_{\nu,\theta}$) when $\theta = \theta_0$ and also $a_0 := a^\delta_{b_0,\theta_0}$, and $J_0(\cdot) := J_{\delta,b_0,\theta_0}(\cdot)$.

3. Some asymptotic properties of the CGEM estimating equation

In this section, we established some consistency and identifiability properties enjoyed by the CGEM estimating equation in θ. They are merely encouraging preliminary results. We expect to give more complete results elsewhere. To simplify

our statement, we consider, the normalized version $\mathbf{y}^T A_{b,\theta}(I - A_{b,\theta})\mathbf{y}/\mathrm{tr}A_{b,\theta} = 1$ which is, of course, an equivalent estimating-equation.

For any positive $\delta, b, \theta, b_0, \theta_0$, we can define the following weighted mean:

$$\psi(\delta, b, \theta, b_0, \theta_0) := \frac{1}{\int_{-\pi}^{\pi} a_{b,\theta}^{\delta}(\lambda)d\lambda} \int_{-\pi}^{\pi} [a_{b,\theta}^{\delta}(\lambda)]^2 \left(\frac{b_0 g_{\nu,\theta_0}^{\delta}(\lambda)}{b g_{\nu,\theta}^{\delta}(\lambda)} - 1 \right) d\lambda.$$

Theorem 3.1. *For any fixed b, θ, we have the following convergence in probability:*

$$\frac{\mathbf{y}^T A_{b,\theta}(I - A_{b,\theta})\mathbf{y}}{\mathrm{tr}A_{b,\theta}} - 1 \to \psi(\delta, b, \theta, b_0, \theta_0) \quad \text{as} \quad n \to \infty,$$

and ψ has a very simple small-δ equivalent:

$$\psi(\delta, b, \theta, b_0, \theta_0) \to 2\nu(2\nu + 1)^{-1} \left(\frac{b_0 {\theta_0}^{2\nu}}{b\theta^{2\nu}} - 1 \right) \quad \text{as } \delta \to 0.$$

The first part of this theorem is in fact not restricted to the Matérn family; this is a rather direct consequence of universal results of time-series theory about quadratic forms constructed from (product of power, possibly negative, of) Toeplitz matrices. In order to be allowed to apply such classical results (for instance, those stated by Azencott and Dacunha-Castelle (1986) and used in their analysis of maximum likelihood estimation, or its Whittle approximation), it is sufficient that $(\theta, \lambda) \mapsto g_{\nu,\theta}^{\delta}(\lambda)$ be three-times continuously differentiable on $\Theta \times [-\pi, \pi]$ where Θ is a compact interval not containing 0; and this can be checked by applying the classical Weierstrass M-test.

The second part uses several appromixations to integrals which can be obtained by the technics of Section 5.

Theoreme 2.1 thus says that the limit of the normalized CGEM equation has a small-δ equivalent which is a very simple monotonic function of $b\theta^{2\nu}$. Clearly if b is fixed at any value b_1, then the unique root θ_1 of this large-n-small-δ equivalent equation will satisfy $b_1\theta_1^{2\nu} = b_0\theta_0^{2\nu}$. This gives support to the first of the two underlying ideas given in the introduction.

4. Mean square error inefficiencies of CGEM-EV to ML for the variance and range parameters

Let $(\hat{b}_{\mathrm{ML}}, \hat{\theta}_{\mathrm{ML}})$ be a maximizer of the likelihood function on $B \times \Theta$ where B and Θ are compact intervals not containing 0 and such that (b_0, θ_0) is in the

interior of $B \times \Theta$. Then, since the classical identifiability and regularity conditions are well fulfilled, for any fixed $\delta > 0$, it is a well known result of times-series theory (Azencott and Dacunha-Castelle (1986)) that $(\hat{b}_{\rm ML}, \hat{\theta}_{\rm ML})$ is a.s. consistent and satisfy:

$$n^{1/2}\left(\begin{bmatrix} \hat{b}_{\rm ML} \\ \hat{\theta}_{\rm ML} \end{bmatrix} - \begin{bmatrix} b_0 \\ \theta_0 \end{bmatrix}\right) \to_{\mathcal{D}} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, 4\pi \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}\right)$$

where

$$\begin{bmatrix} \sigma_1^2 \\ \sigma_{12} \\ \sigma_2^2 \end{bmatrix} = \left(\int a_0^2 h_0\right)^{-2} J_0(h_0)^{-1} \begin{bmatrix} b_0^2 \int a_0^2 h_0^2 \\ -b_0 \int a_0^2 h_0 \\ \int a_0^2 \end{bmatrix}.$$

Note that the factor $(\int a_0^2 h_0)^{-2} J_0(h_0)^{-1}$ is the determinant (> 0 since $h_0(\cdot)$ cannot be a constant function) of the information matrix and has this form only if $\int a_0^2 h_0 \neq 0$. And the same regularity conditions on our time-series Matérn model $f^{\delta}_{\nu,b,\theta}$ are also sufficient to show (by the usual Taylor series argument) the first part of the following theorem:

Theorem 4.1. *Let $(\hat{b}_{\rm EV}, \hat{\theta}_{\rm GEEV})$ be a consistent root of the CGEM-EV estimating equation (1.3). If $\int a_0^2 h_0 \neq 0$ then*

$$n^{1/2}\left(\begin{bmatrix} \hat{b}_{\rm EV} \\ \hat{\theta}_{\rm GEEV} \end{bmatrix} - \begin{bmatrix} b_0 \\ \theta_0 \end{bmatrix}\right) \to_{\mathcal{D}} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, 4\pi \begin{bmatrix} v_1 & v_{12} \\ v_{12} & v_2 \end{bmatrix}\right)$$

where

$$\left\{\begin{array}{lcl} v_1 & = & b_0^2 \int a_0^{-2} g_0^2 \\ v_{12} & = & -b_0 c_0 \left(\int a_0^2 h_0\right)^{-1} \\ v_2 & = & c_0 \int a_0^2 \left(\int a_0^2 h_0\right)^{-2} \end{array}\right., \quad \text{and} \quad c_0 = J_0(g_0/a_0^2).$$

When $\delta \to 0$ we have $\int a_0^2 h_0 / \int a_0^2 \to 2\nu/\theta_0$; and defining by $I^1_{\delta,b_0,\theta_0} := v_1/\sigma_1^2$ (resp. $I^2_{\delta,b_0,\theta_0} := v_2/\sigma_2^2$) the asymptotic mean square error inefficiency of CGEM-EV to ML for b_0 (resp. for θ_0), these 2 inefficiencies have the following common limit

$$I^i_{\delta,b_0,\theta_0} \to \frac{\sqrt{\pi}}{2}\left(\frac{2\nu+1}{2\nu}\right)^2 \frac{\Gamma\left(\nu+\frac{1}{2}\right)^2 \Gamma\left(2\nu+\frac{1}{2}\right)}{\Gamma(\nu)^2\Gamma(2\nu+1)} \quad \text{as} \quad \delta \to 0, \quad \text{for} \quad i \in \{1,2\}.$$

The first rather surprising fact is that this large-n-small-δ inefficiency is not function of the underlying b_0, nor of θ_0. Secondly, Table 4.1 clearly demonstrates that the CGEM-EV estimates are asymptotically nearly efficient provided ν is not too large. Asymptotic full-efficiency is reached for ν close to 1/2.

ν	1/2	1	3/2	2	5/2	3	7/2	4
$\lim_{\delta\to 0} I^i_{\delta,b_0,\theta_0}$	1	1.04093	$\frac{10}{9}$	1.18596	$\frac{63}{50}$	1.33174	$\frac{1716}{1225}$	1.46727

Table 4.1. Limit of asymptotic MSE inefficiencies $I^i_{\delta,b_0,\theta_0}$, $i \in \{1, 2\}$, for some "typical" values of ν

Of course it would be interesting to study how fine and large the grid must be in order that the CGEM-EV statistical performances be comparable to ML. One may guess that the signal-to-noise ratio must also be "not too small", even if its value does not matter asymptotically. Since the main improvements in numerical performance occur for multidimensional process (with randomized-traces used instead of the exact traces in (1.3)), a first Monte Carlo study has been done for two-dimensional random fields in Girard (2009).

This approch may be extended in a simple way to the case of unknown variance, says σ_ε^2, of the measurement error (or unknown nugget effect). Indeed it is well known (Stein (1999, Section 6.2)) that the simple average of local squared differences like $(y_i - y_{i-1})^2$ would yield here a consistent estimate $\hat\sigma_\varepsilon^2$, and this is easily generalized to multidimensional processes which are mean square continuous. Notice that the Bayesian filter is still $A_{b,\theta}$ except that b now denotes the signal-to-noise ratio s.t. $\mathrm{var}(Z) = b\sigma_\varepsilon^2$. Then the whole CGEM-EV approach can be applied after having replaced $\mathbf{y}$ by its "standardized" version $\mathbf{y}/\hat\sigma_\varepsilon$. It would be thus very interesting to study whether similar near-efficiency results still hold or if a more elaborate estimate for σ_ε is required.

Concerning the case with known noise *and* signal variances, even if the assumption that b_0 were known is rather restrictive, it may be worth to point out the following very neat result: denoting by $\hat\theta_{\mathrm{ML}_0}$ the ML estimate with b fixed at b_0, its asymptotic behavior is classically obtained

$$n^{1/2}\left(\hat\theta_{\mathrm{ML}_0} - \theta_0\right) \to_{\mathcal{D}} N\left(0, 4\pi\left(\int a_0^2 h_0^2\right)^{-1}\right),$$

and denoting by $\hat\theta_{\mathrm{GE}_0}$ a consistent solution of the estimating-equation (1.3) with now b fixed at b_0, we have

Theorem 4.2. *If* $\int a_0^2 h_0 \neq 0$ *then*

$$n^{1/2}\left(\hat\theta_{\mathrm{GE}_0} - \theta_0\right) \to_{\mathcal{D}} N\left(0, 4\pi\int a_0^2\left(\int a_0^2 h_0\right)^{-2}\right)$$

and the ratio of this asymptotic variance over the one of $\hat\theta_{\mathrm{ML}_0}$*, says* $I^0_{\delta,b_0,\theta_0}$*, satisfies:*

$$I^0_{\delta,b_0,\theta_0} = 1 + J_0(h_0) \to 1 \quad \text{as} \quad \delta \to 0.$$

Asymptotic (i.e. large-n-small-δ) full-efficiency is thus now enjoyed by the CGEM estimating function for *any* value of $\nu \geq 1/2$.

To finish, let us comment on the somewhat more natural $\hat{b}_{\mathrm{EV}}$-based alternative to ML which would be akin to the hybrid estimate proposed by Zhang and Zimmerman (2007) : choose $\hat{\theta}_{\mathrm{H}}$ so as to maximize the likelihood function with $\hat{b}_{\mathrm{EV}}$ plugged-in. As to the computational aspects, $\hat{\theta}_{\mathrm{H}}$ may be much less easily computed than $\hat{\theta}_{\mathrm{GEEV}}$ (for example when an iterative solver is used for $A_{b,\theta}\mathbf{y}$) since it requires a log-determinant instead of a trace. So it is worth to point out that again a neat result (which can be similarly obtained by the technics used to prove Theorem 4.1) holds true here: the large-n mean square error of $\hat{\theta}_{\mathrm{H}}$ over the one of $\hat{\theta}_{\mathrm{GEEV}}$ tends to 1 as $\delta \to 0$.

5. Proofs

In the following, c denotes a constant that may change from line to line (that is, c will only possibly depend on ν and on the positive lower and upper bounds for the candidate b's and those for θ (says $\underline{b}, \bar{b}, \underline{\theta}, \bar{\theta}$)). And $\mathrm{O}(\delta^{2\nu})$ denotes a term whose absolute value is bounded by $c\,\delta^{2\nu}$. Without loss of generality we assume that $\delta \leq 1$.

Firstly, it is easily checked from the definition of $g^*_{\nu,\alpha}(\lambda)$, that the un-aliased version of $h^{\delta}_{\nu,\theta}$ has a very simple form:

$$h^*_{\nu,\alpha}(\lambda) = \alpha^{-1}\left(2\nu - (2\nu+1)G\left(\frac{\lambda}{\alpha}\right)\right), \quad \text{where} \quad G(\omega) = \frac{1}{1+\omega^2}. \tag{5.1}$$

We first collect intermediary results in a few lemmas which might be of interest for other studies of this small-δ Matérn time-series model.

Lemma 5.1. *We have*

$$c\,\delta^{2\nu} \leq g^*_{\nu,\delta\theta}(\lambda) \leqslant g^{\delta}_{\nu,\theta}(\lambda) \leqslant g^*_{\nu,\delta\theta}(\lambda) + \mathrm{O}(\delta^{2\nu}),$$

and

$$g^*_{\nu,\delta\theta}(\lambda)\left|h^{\delta}_{\nu,\theta}(\lambda) - \delta\, h^*_{\nu,\delta\theta}(\lambda)\right| = \mathrm{O}(\delta^{2\nu}).$$

Lemma 5.2. *We have*

$$a^*_{b,\delta\theta}(\lambda) \leqslant a^{\delta}_{b,\theta}(\lambda) \leqslant a^*_{b,\delta\theta}(\lambda) + \mathrm{O}(\delta^{2\nu}),$$

and

$$a^{\delta}_{b,\theta}(\lambda) \leqslant c\, g^{*}_{\nu,b\theta}(\lambda).$$

Proof of these Lemmas. From their definition, we have $g_{\nu,\alpha}(\lambda) - g^{*}_{\nu,\alpha}(\lambda) = \sum_{k\neq 0} g^{*}_{\nu,\alpha}(\lambda + 2\pi k)$. For any $\lambda \in [-\pi,\pi]$, for any $k \geq 1$, the monotonicity of $g^{*}_{\nu,\alpha}$ implies $g^{*}_{\nu,\alpha}(\lambda + 2\pi k) \leq g^{*}_{\nu,\alpha}(2\pi(k-1/2)) \leq c\,\alpha^{2\nu}/(k-1/2)^{2\nu+1}$; and summing these terms over $k = 1, 2, \cdots$, thus gives a term $\mathrm{O}(\alpha^{2\nu})$. This is shown similarly for the sum over $k = -1, -2, \cdots$. We now prove the second part of Lemma 5.1. Omitting the index ν and denoting $\dot{g}_{\alpha} := \partial g_{\alpha}/\partial\alpha$, $\dot{g}^{*}_{\alpha} := \partial g^{*}_{\alpha}/\partial\alpha$, one can decompose $\alpha g^{*}_{\alpha}\left(\delta^{-1}h^{\delta}_{\theta} - h^{*}_{\alpha}\right) = \alpha g^{*}_{\alpha}\left(\dot{g}_{\alpha}/g_{\alpha} - \dot{g}^{*}_{\alpha}/g^{*}_{\alpha}\right) = \left(g^{*}_{\alpha} - g_{\alpha}\right)\left(\alpha\dot{g}_{\alpha}/g_{\alpha}\right) + \alpha\left(\dot{g}_{\alpha} - \dot{g}^{*}_{\alpha}\right)$. Observing that $\alpha\dot{g}^{*}_{\alpha} = 2\nu g^{*}_{\nu,\alpha} + c\,g^{*}_{\nu+1,\alpha}$ yields $\alpha\left(\dot{g}_{\alpha} - \dot{g}^{*}_{\alpha}\right) = \mathrm{O}(\alpha^{2\nu})$. Lastly, we can bound $\alpha\left|\dot{g}_{\alpha}/g_{\alpha}\right| \leq \alpha\left|\dot{g}_{\alpha}/g^{*}_{\alpha}\right| \leq \alpha\left|\dot{g}^{*}_{\alpha}/g^{*}_{\alpha}\right| + \alpha\left|\dot{g}_{\alpha} - \dot{g}^{*}_{\alpha}\right|/g^{*}_{\alpha} \leq \mathrm{O}(1) + \alpha\left|\dot{g}_{\alpha} - \dot{g}^{*}_{\alpha}\right|/c\alpha^{2\nu}$ from equation (5.1) and the first inequality of Lemma 5.1.

Lemma 5.3. *If $\nu > 1/2$, then for any continuous function $F : \mathbb{R} \to \mathbb{R}_{+}$ such that $0 < \int_{-\infty}^{\infty} F(x)\mathrm{d}x < +\infty$, we have, as $\delta \to 0$*

$$\int a^{\delta}_{b,\theta}(\lambda) F\left(\frac{\lambda}{\delta\theta}\right)\mathrm{d}\lambda \sim \delta\theta \int_{-\infty}^{\infty} F(x)\mathrm{d}x, \quad \int \left[a^{\delta}_{b,\theta}(\lambda)\right]^{2} F\left(\frac{\lambda}{\delta\theta}\right)\mathrm{d}\lambda \sim \delta\theta \int_{-\infty}^{\infty} F(x)\mathrm{d}x.$$

Proof. The first part of Lemma 5.2 and the boundedness of F permits us to replace $a^{\delta}_{b,\theta}$ by the much more manageable $a^{*}_{b,\alpha}$ (where $\alpha = \delta\theta$) with an error $\mathrm{O}(\delta^{2\nu})$ which is enough when $\nu > 1/2$. Letting $\bar{\lambda} = \alpha^{\epsilon+(2\nu/(2\nu+1))}$ where ϵ is any arbitrarily small constant > 0, we see from the monotonicity of $g^{*}_{\nu,\alpha}$, that $\inf_{\lambda\in[-\bar{\lambda},\bar{\lambda}]} a^{*}_{b,\alpha}(\lambda) \geq \left(1 + (2\pi/\underline{b})\left[g^{*}_{\nu,\alpha}(\bar{\lambda})\right]^{-1}\right)^{-1}$ which tends to 1 since $g^{*}_{\nu,\alpha}(\bar{\lambda}) \geq \alpha^{2\nu}/\left(2\bar{\lambda}^{2}\right)^{\nu+1/2}$ (from α small enough) which tends to $+\infty$. And the same is true for $\left[a^{*}_{b,\alpha}(\lambda)\right]^{2}$. It remains to observe that both $\int_{-\pi}^{\pi} F\left(\frac{\lambda}{\delta\theta}\right)\mathrm{d}\lambda$ and $\int_{-\bar{\lambda}}^{\bar{\lambda}} F\left(\frac{\lambda}{\delta\theta}\right)\mathrm{d}\lambda$ are equivalent to the claimed term.

Lemma 5.4. *We have*

$$\int a^{\delta}_{b,\theta}(\lambda)\mathrm{d}\lambda \sim (\delta\theta)^{\frac{2\nu}{2\nu+1}}(2\pi c_{\nu} b)^{\frac{1}{2\nu+1}} \int_{-\infty}^{\infty} (1 + x^{2\nu+1})^{-1}\mathrm{d}x,$$
$$\int \left[a^{\delta}_{b,\theta}(\lambda)\right]^{2}\mathrm{d}\lambda \sim (\delta\theta)^{\frac{2\nu}{2\nu+1}}(2\pi c_{\nu} b)^{\frac{1}{2\nu+1}} \int_{-\infty}^{\infty} (1 + x^{2\nu+1})^{-2}\mathrm{d}x.$$

Proof. By Lemma 5.2 we can replace, with enough accuracy, the filter by its unaliased version in these integrals. Next the claimed equivalents can be obtained by

a change of variable $s = \lambda/\alpha^{\frac{2\nu}{2\nu+1}}$ and an application of the dominated convergence theorem.

We can now outline the proofs of Theorem 3.1 and Theorem 4.1, beginning by the second one.

Proof of Theorem 4.1. i) First it is directly seen that $v_2/\sigma_2^2 = J_0(h_0)c_0$. Let $w(\lambda) := \left[a^\delta_{b,\theta}(\lambda)\right]^2$. Consider first the constant c_0. Since $1+J_{\delta,b,\theta}(g^\delta_{\nu,\theta}/w)$ is the ratio of $\int w \int \left[g^\delta_{\nu,\theta}(\lambda)\right]^2/w(\lambda)\mathrm{d}\lambda$, where $\left[g^\delta_{\nu,\theta}\right]^2/w = \left[g^\delta_{\nu,\theta} + (2\pi b)^{-1}\right]^2$, over $\left[\int g^\delta_{\nu,\theta}\mathrm{d}\lambda\right]^2$ ($=1$), it suffices to observe that

$$\left|\int {g_{\nu,\alpha}}^2 - \int {g^*_{\nu,\alpha}}^2\right| \leq \int \left|{g_{\nu,\alpha}}^2 - {g^*_{\nu,\alpha}}^2\right| \leq c\alpha^{2\nu}\int (g_{\nu,\alpha} + g^*_{\nu,\alpha}) \leq 2c\alpha^{2\nu}$$

and that $\int {g^*_{\nu,\alpha}}^2 \sim \alpha^{-1}c_\nu^2 \int_{-\infty}^{\infty}(1+x^2)^{-2\nu-1}\mathrm{d}x$ to see that this is also the dominant term of $J_{\delta,b,\theta}(g^\delta_{\nu,\theta}/w)/\int w$. Secondly, the equivalent of $J_0(h_0)$ can be obtained as follows. The numerator of $J_{\delta,b,\theta}(h^\delta_{\nu,\theta})$ is

$$\frac{1}{\int w}\int w(h^\delta_{\nu,\theta})^2 - \left|\frac{1}{\int w}\int w(h^\delta_{\nu,\theta})\right|^2 = \frac{1}{\int w}\int w(h^\delta_{\nu,\theta} - 2\nu/\theta)^2 - \left|\frac{1}{\int w}\int w(h^\delta_{\nu,\theta} - 2\nu/\theta)\right|^2.$$

If we take for granted that the un-aliased $\delta h^*_{\nu,\delta\theta}$ can be substituted for $h^\delta_{\nu,\theta}$, then the numerator is simplified using equation (5.1), and its equivalent is deduced from Lemmas 5.3 (case $\nu > 1/2$) and 5.4 (with $G_\alpha(\lambda) := G(\lambda/\alpha)$ defined from (5.1)) as follows:

$$\frac{1}{\int w}\int w\left|(2\nu+1)G_\alpha/\theta\right|^2 - \left|\frac{1}{\int w}\int w(2\nu+1)G_\alpha/\theta\right|^2 \sim ((2\nu+1)/\theta)^2\frac{1}{\int w}\delta\theta\int_{-\infty}^{\infty}G(x)^2\mathrm{d}x.$$

Note that this can also be proved for $\nu = 1/2$ by using the closed expression of w available in this case. On the other hand, the denominator, also for the un-aliased version, is $\left|\left(\int w\right)^{-1}\int w(\lambda)[2\nu - (2\nu+1)G(\lambda/\alpha)]/\theta\right|^2$ which tends to $(2\nu/\theta)^2$ again by Lemmas 5.3 and 5.4. This proves the claimed result for the θ-inefficiency provided the un-aliased substitution was authorized. To see this, e.g. for $\int w(h^\delta_{\nu,\theta} - 2\nu/\theta)^2$, it suffices, to decompose $\left(h^\delta_{\nu,\theta} - 2\nu/\theta\right)^2 - \left(\delta h^*_{\nu,\delta\theta} - 2\nu/\theta\right)^2 = \left(h^\delta_{\nu,\theta} - \delta h^*_{\nu,\delta\theta}\right)^2 + 2\left(\delta h^*_{\nu,\delta\theta} - 2\nu/\theta\right)\left(h^\delta_{\nu,\theta} - \delta h^*_{\nu,\delta\theta}\right)$ and to bound from Lemmas 5.1-5.3, for example for the second term:

$$\begin{aligned}\int w\left|h^\delta_{\nu,\theta} - \delta h^*_{\nu,\delta\theta}\right|G_\alpha &\leq c\int w^{1/2}g^*_{\nu,\delta\theta}\left|h^\delta_{\nu,\theta} - \delta h^*_{\nu,\delta\theta}\right|G_\alpha \\ &\leq c\,\mathrm{O}(\delta^{2\nu})\int w^{1/2}G_\alpha = \mathrm{O}(\delta^{2\nu+1}).\end{aligned}$$

ii) By a simple algebraic manipulation (and using that $\int g_0 = 1$) we can write

$$\frac{v_1}{\sigma_1^2} = \frac{v_2}{\sigma_2^2}\big(1 + c_0^{-1}\big)\frac{1}{1 + J_0(h_0)}.$$

(as a side note, this implies $I^1_{\delta,b_0,\theta_0} < I^2_{\delta,b_0,\theta_0}$ since, of course, $J_0(h_0)c_0 > 1$ by a Cauchy-Schwarz type inequality). Thus observing that $\lim_{\delta\to 0} c_0 = +\infty$ and $\lim_{\delta\to 0} J_0(h_0) = 0$ were, in fact, intermediary results of the paragraph i) above, we obtain the claimed result for the b-inefficiency.

Proof of Theorem 3.1. Again let $w(\lambda) := \big[a^\delta_{b,\theta}(\lambda)\big]^2$ and let $g := g^\delta_{\nu,\theta}$. We easily see, from the expression of the equivalents of $\int w^{1/2}$ and $\int w$ of Lemma 5.4, that it is sufficient to prove that $\int w(g_0/g) \sim (\theta_0^{2\nu}/\theta^{2\nu})\int w$. Let us first study this integral with the unaliased versions $g^* := g^*_{\nu,\delta\theta}$ and $g_0^* := g^*_{\nu,\delta\theta_0}$ in place of g and g_0. Starting from the definition (1.1) of g^*, we can write

$$\theta^{2\nu} g_0^*(\lambda)/(\theta_0^{2\nu} g^*(\lambda)) = \big(\theta^2 + (\lambda/\delta)^2\big)^{\nu+1/2}\big(\theta_0^2 + (\lambda/\delta)^2\big)^{-(\nu+1/2)} = 1 + F(\lambda/\delta)$$

with

$$F(\lambda) := \big(1 + (\theta^2 - \theta_0^2)/(\theta_0^2 + \lambda^2)\big)^{\nu+1/2} - 1.$$

Note that F is integrable over $\mathbb{R}$ and has the same sign as $\theta^2 - \theta_0^2$. By Lemma 5.3, if $\nu > 1/2$ then we obtain $\int w(\lambda)F(\lambda/\delta)\mathrm{d}\lambda = \mathrm{o}(\int w)$ and thus $\int w(\theta^{2\nu} g_0^*/(\theta_0^{2\nu} g^*)) \sim \int w$. The same result for the case $\nu = 1/2$ can be easily obtained (from the closed expression of w in this case). It remains to bound the effect of the aliasing:

$$\begin{aligned}\left|\int w(g_0^*/g^*) - \int w(g_0/g)\right| &\le \int w\,|(g_0^* g - g_0 g^*)/g^* g| \\ &\le c\int w^{1/2}\,|(g_0^* g - g_0 g^*)/g^*| \\ &\le c\int w^{1/2}\,|g_0^* - g_0| + c\int w^{1/2}\,|(g - g^*)g_0^*/g^*|\end{aligned}$$

where the second inequality results from the obvious bound $w^{1/2}/bg \le 1$. Now the first term of this sum is clearly $\mathrm{O}(\delta^{2\nu})$ from Lemma 5.1; this also holds for the second because of the boundedness of g_0^*/g^* which results from that of $F(\lambda/\delta)$.

REFERENCES

Azencott, R., Dacunha-Castelle, D., 1986. Séries d'Observations Irrégulières. Masson.

Chen, W., Hurvich, C.M., and Lu, Y., 2006. On the correlation matrix of the discrete Fourier transform and the fast solution of large Toeplitz systems for long-memory time series. J. Amer. Statist. Assoc. 101(474), 812822.

Du, J., Zhang, H., Mandrekar, V., 2009. Fixed-domain asymptotic properties of tapered maximum likelihood estimators. Ann. Statist. 37, 3330-3361.

Girard, D.A., 2009. A fast, near efficient, randomized-trace based method for fitting stationary Gaussian spatial models to large noisy data sets in the case of a single range-parameter. Unpublished.

Kessler, M., 2000. Simple and explicit estimating functions for a discretely observed diffusion process. Scand. J. Statist. 27, 65-82.

Rasmussen, C.E., Williams, C.K.I., 2006. Gaussian Processes for Machine Learning. The MIT Press.

Stein, M.L., 1999. Interpolation of Spatial Data: Some Theory for Kriging. Springer.

Yadrenko, M.I., 1983. Spectral Theory of Random Fields. Optimization Software, Inc.

Zhang, H., 2004. Inconsistent estimation and asymptotically equivalent interpolation in model-based geostatistics. J. Amer. Statist. Assoc. 99, 250-261.

Zhang, H., Zimmerman, D.L., 2007. Hybrid estimation of semivariogram parameters. Mathematical Geology 39 (2), 247-260.